

Full Paper

# Bending of DNA duplexes with mutation motifs

Michal Ružička<sup>1,2</sup>, Přemysl Souček<sup>1</sup>, Petr Kulhánek<sup>1,3</sup>, Lenka Radová<sup>1</sup>,  
Lenka Fajkusová<sup>4</sup>, and Kamila Réblová<sup>1\*</sup>

<sup>1</sup>CEITEC - Central European Institute of Technology, Masaryk University, 62500 Brno, Czech Republic, <sup>2</sup>Department of Condensed Matter Physics, Faculty of Science, Masaryk University, 61137 Brno, Czech Republic, <sup>3</sup>National Centre for Biomolecular Research, Faculty of Science, Masaryk University, 62500 Brno, Czech Republic, and <sup>4</sup>Centre of Molecular Biology and Gene Therapy, University Hospital Brno, 62500 Brno, Czech Republic

\*To whom correspondence should be addressed. Tel. +420 549497586. Email: kristina@physics.muni.cz

Edited by Dr Mikita Suyama

Received 18 January 2019; Editorial decision 23 May 2019; Accepted 27 May 2019

## Abstract

Mutations can be induced by environmental factors but also arise spontaneously during DNA replication or due to deamination of methylated cytosines at CpG dinucleotides. Sites where mutations occur with higher frequency than would be expected by chance are termed hotspots while sites that contain mutations rarely are termed coldspots. Mutations are permanently scanned and repaired by repair systems. Among them, the mismatch repair targets base pair mismatches, which are discriminated from canonical base pairs by probing altered elasticity of DNA. Using biased molecular dynamics simulations, we investigated the elasticity of coldspots and hotspots motifs detected in human genes associated with inherited disorders, and also of motifs with Czech population hotspots and *de novo* mutations. Main attention was paid to mutations leading to G/T and A+/C pairs. We observed that hotspots without CpG/CpHpG sequences are less flexible than coldspots, which indicates that flexible sequences are more effectively repaired. In contrary, hotspots with CpG/CpHpG sequences exhibited increased flexibility as coldspots. Their mutability is more likely related to spontaneous deamination of methylated cytosines leading to C > T mutations, which are primarily targeted by base excision repair. We corroborated conclusions based on computer simulations by measuring melting curves of hotspots and coldspots containing G/T mismatch.

**Key words:** DNA bending, Muts protein, mutations, hotspots–coldspots, free energy calculations

## 1. Introduction

Mutations in DNA are responsible for various inherited disorders and cancer, but they are also crucial for evolution. Here we focus on elastic properties of DNA sequences with hereditary mutations, which can be obtained from parents (from a mother via oocyte or from a father through sperm) or they arise *de novo* (during formation of parent's germ cells or post-zygotically).<sup>1,2</sup> Mutations can be induced by environmental factors, but they also arise spontaneously: (i) due to deamination of the methylated cytosine (mC),<sup>3,4</sup> (ii) during

DNA replication.<sup>5</sup> With a fully functional repair system, the frequency of spontaneous error is ca 1 per 10<sup>9</sup>–10<sup>10</sup> base pairs per replication.<sup>6</sup> Mutations are not equally distributed along the DNA sequence.<sup>7,8</sup> Sites that possess higher mutation frequency than would be expected by chance are considered as hotspots, while sequences with lower frequencies than would be expected by chance are coldspots. The CpG dinucleotide is a well-known hotspot where mC converts to thymine resulting in TpG (or CpA on the other strand) transitions.<sup>9</sup> Hypermutability of CpG reflects spontaneous

methylation of cytosine, which is a common post-replicative modification of DNA in both bacteria and eukaryotes.<sup>10</sup> These errors are recognized by thymine DNA glycosylases within the base excision repair (BER) pathway where the glycosylic bond between the thymine base and the ribose is cleaved, which triggers downstream events of the BER.<sup>10,11</sup> Despite the effective repair, this mutation is very frequent.<sup>12–14</sup> As a result, CpG sites occur less frequently in the genome, than would be expected by chance.<sup>15,16</sup> Except for CpG dinucleotide, CpHpG motif (where H stands for A, C or T) was also detected as a hotspot.<sup>17</sup> Another example is GTAAGT sequence, which is mainly associated with insertion/deletion errors.<sup>18</sup> It was observed that a sequence of  $\pm 2$  nucleotides (nt) around a mismatch affects relative rates of single nucleotide variations and that mismatches are not repaired with the same efficiency,<sup>19–21</sup> which highlights the importance of both the mismatch type and the neighbouring nucleotide sequence. Mismatches and small insertion/deletion loops are targeted by MutS protein (homologue of eukaryotic MutSz) within mismatch repair (MMR). Two subunits (MSH2 and MSH6) form human MutSz,<sup>6</sup> which moves along DNA and searches for mismatches.<sup>22,23</sup> Mismatch detection is coupled with conformational changes of the subunits that create a transient clamp around DNA.<sup>24</sup> When the MutSz identifies a mismatch, it provokes ATP binding that induces a sliding clamp with high stability. MutL protein is then recruited, which initiates the repair process.<sup>25</sup> The clamp relaxation is associated with ATP hydrolysis, thus the MutSz switches between open/closed states.<sup>26</sup> Crystallographic structures of MutS/DNA (or MutSz/DNA) complexes revealed that DNA with a mismatch is bent by ca 60°.<sup>27–30</sup> A model where MutSz slides along DNA and scans for flexible regions corresponding to mismatches has been proposed.<sup>22</sup> The mismatch is further recognized by phenylalanine 432 and glutamine 434 from conserved Phe-X-Glu motif in MutSz. The phenylalanine intercalates into minor groove and stacks onto mismatched base while the glutamine creates a hydrogen bond with it.<sup>30</sup>

Molecular dynamics (MDs) is a computational method that can describe the behaviour of molecules at the atomic level on nanosecond (ns) to microsecond ( $\mu$ s) time scale. Although this method utilizes simplified potential energy functions and fixed atomic charges it has provided valuable information about DNA flexibility.<sup>31–34</sup> Bending of DNA with different mismatches (in the absence of protein) was analysed by Sharma et al.<sup>35</sup> where a biased MD method (umbrella sampling) was used because DNA cannot bend spontaneously to the conformation observed in the complex with MutS protein. Analysis of mismatches in one sequence context revealed different bending free energies which partially correlated with the binding affinities of MutS to the mismatches. The umbrella sampling method was also used in the study of bending of various B-DNA sequences (AT, GC and A-tract sequences) containing G/T, G/A and C/C mismatches and abasic sites.<sup>36,37</sup> Structural analysis of DNA mismatches in three different sequence environments was also carried using unrestrained MD simulations.<sup>38</sup> The study showed to which extent different mismatches induce alterations of the local structure and which base pair arrangements are most populated. The relation between DNA mutability and flexibility was however not reported so far.

In our previous study,<sup>39</sup> we analysed the presence of germline mutations (obtained from the Human Gene Mutation Database—HGMD, <http://www.hgmd.cf.ac.uk/ac/index.php>) in the 5-nt DNA segments repeatedly occurring in five genes (*PAH*, *LDLR*, *CFTR*, *F8* and *F9*) associated with common inherited disorders. This approach contrasts with the conventional strategy because hotspots are usually

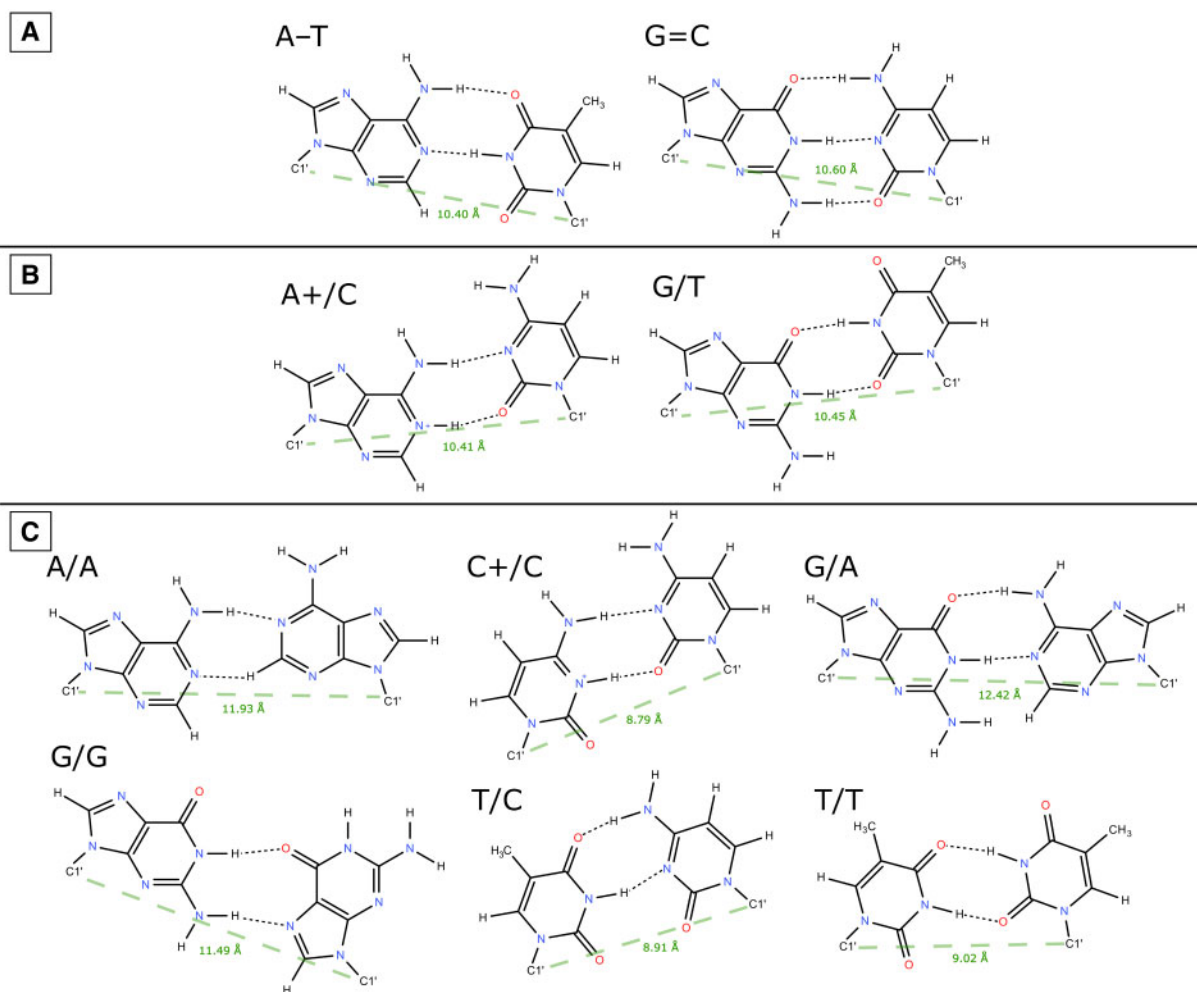
derived from a mutation spectrum. The occurrence of mutations in the mutation spectrum, however, often reflects particular population where an effect of a common ancestor plays a role (founder effect), i.e. frequent mutation originate from a single mutation event and is spread throughout the population as observed in our<sup>40</sup> and other studies.<sup>41,42</sup> We detected statistically significant coldspots and hotspots and determined characteristic sequence patterns for each group, particularly, coldspots contained purine tract while hotspots showed alternating purine-pyrimidine sequence often with the presence of CpG dinucleotide. In the next step, we employed an adaptive biasing force (ABF) MD method<sup>43</sup> to better understand the relation between DNA-bending properties and DNA mutability. In particular, we analysed the global bending properties of two coldspots and two hotspots with a G/T pair and derived the free energy required to bend a straight DNA duplex towards the geometry found in the complex with MutSz protein. We observed that the coldspots are inherently more flexible than the hotspots.

In this study, we extended analysed dataset and derived bending free energies for 10 coldspots and ten hotspots with different mismatches. Primary attention was paid to G/T and A+/C mismatches which are results of transitions occurring more frequently than transversions.<sup>44</sup> In addition, G/T and A+/C base pair geometries are near-isosteric to canonical pairs. We also analysed other mismatches (A/A, G/G, A/G, C/T, C/C and T/T), but for the symmetrical G/G, A/A and C/C mismatches, we did not get stable base pairs during bending. Moreover, we carried out a bending analysis of 10 Czech population hotspots and 15 *de novo* mutations with G/T and A+/C pairs. The computations are complemented by experimental melting curves measured for eight coldspots and eight hotspots with G/T mismatch.

## 2. Materials and methods

### 2.1. The building of DNA 15-mers with mismatches

The 15-nt long DNA duplexes with central mismatch pair were built as a right-handed B-DNA by the nab module of AMBER 14.<sup>45</sup> Terminal 5-nt segments (underlined) were identical for all systems, only the central 5-nt region varied according to the analysed motif, i.e. forward sequence was 5'GAACC XXXXX CTAGG3', the reverse sequence was 5'CCTAG XXXXX GGTTC3', the position of the mismatch is in bold. To maintain consistency during bending analysis when we analysed purine/pyrimidine mismatches, the purine base and the surrounding  $\pm 2$  nt sequence was always put into the forward sequence. Our tests showed that the placement into the reverse strand has negligible impact on the resulting bending free energy. DNA was then solvated with TIP3P water molecules<sup>46</sup> in a truncated octahedral periodic box and neutralized with sodium ions<sup>47</sup> using the tleap module of AMBER, a new version of DNA force field parmbsc1<sup>48</sup> was used. Similarly to the study,<sup>35</sup> DNA mismatched base pairs (G/T, A/C, A/G, G/A, C/T, T/C, C/C, T/T and A/A) were built in *anti-anti* form which is prevailing geometry.<sup>38</sup> Figure 1 shows all analysed mismatch geometries, indicated C1'–C1' distances for A-T and G = C pairs were taken from the previous study<sup>49</sup> while C1'–C1' distances of mismatched base pairs are average values calculated from 150 ns production simulations described below. In the *anti-anti* form, both bases have glycosidic bonds in *anti*-conformation and mutually interact through Watson-Crick (WC) edge (there are three edges: WC, Hoogsteen (H) and Sugar, through which base can interact with another base). The G/G mismatch was built in *anti-syn* geometry where the glycosidic bond of one guanine is rotated about 180° and adopts a *syn* conformation



**Figure 1.** Base pair geometries: (A) Canonical WC base pairs A-T and G = C. (B) Mismatches A+/C and G/T used for calculations of bending free energies. (C) The remaining tested mismatches: A/A, C+/C, G/A, G/G, T/C and T/T.

(Fig. 1), which is considered to be a more stable arrangement.<sup>50,51</sup> In the *anti-syn* G/G mismatch, first base interacts through WC edge, whereas the second one through H edge. Further, A/C and C/C pairs were built protonated. Protonated forms of adenine at N1 and cytosine at N3 positions were taken from the AMBER library. The protonated pairs represent stable arrangements where each pair consists of two H-bonds (Fig. 1). The previous study<sup>35</sup> has shown that protonated A+/C pair is the relevant state for DNA bending.

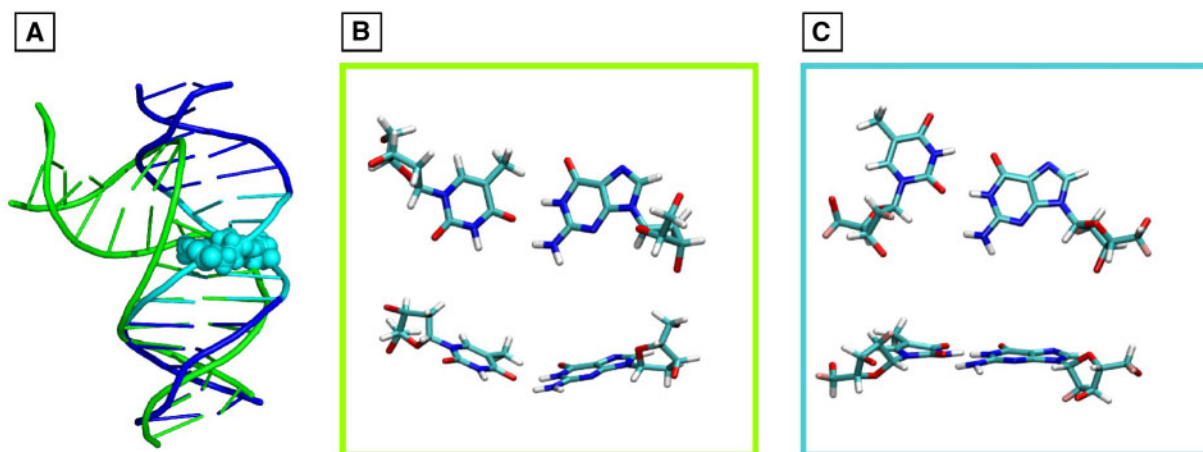
## 2.2. Equilibration and production dynamics

The systems were equilibrated in three steps. First, a minimization of the system was performed in 3,000 steps, then the system was heated to 300 K during 100 ps at a constant volume and finally, a 500 ps long simulation was run at 300 K and 100 kPa at the NPT conditions. Then we ran production dynamics for 150 ns at the NPT conditions. During both equilibration and production dynamics, terminal base pairs were restrained by distance restraints imposed on hydrogen bonds to avoid a formation of flanking bases that could perturb the DNA structure. The built models contained mismatched bases mutually oriented with WC edges close to the optimal base pair arrangement (visualized in Fig. 1). During the minimization/equilibration procedure the mismatched pair spontaneously adopted

the optimal arrangement in the nearest energy minimum (see below where minima are described). Formed mismatches were stable during 150 ns production dynamics. Similarly, we prepared the G/G pair in the *anti-syn* form using xleap module of AMBER; the first guanine was oriented with WC edge and the second one with H edge.

## 2.3. ABF calculations

We used the ABF method<sup>43</sup> to calculate free energies during DNA bending and employed the Multiple Walkers Approach (MWA)<sup>52</sup> to accelerate the calculations. All the ABF simulations were performed in the modified PMEMD programme from AMBER, connected with PMFLib.<sup>53</sup> The length of each ABF simulation was 200 ns (40 ns for each MWA client). Convergence of these 200 ns simulations was tested in our previous study.<sup>39</sup> The collective variable used in the free energy calculations was the mass-weighted root-mean-square distance (RMSD) to a target structure calculated over all of the heavy atoms from terminal 5-nt long segments and the heavy atoms of the backbone from the central 5-nt long segment corresponding to the AMBER mask notation ((: 1-5, 11-20, 26-30) | (:6-10, 21-25@P, OP1, OP2, O3', O5', C3', C4', C5')) & (!@H=). The target structure used in the calculations was bent X-ray DNA duplex from the complex with MutSz protein (PDB ID 2O8B)<sup>30</sup> (Fig. 2A).



**Figure 2.** (A) Superimposed bent X-ray DNA structure (from PDB ID 2O8B) (green) and relaxed MD structure (blue) (structures are superimposed over lower part). In the MD structure, a central 5-nt segment which was either coldspot or hotspot is cyan, and G/T mismatch is highlighted in sphere representation. Detail of G/T pair in the X-ray structure (B) and in MD structure (C).

Further, we imposed a shear wall restraint to keep the mismatch pair stable inside the DNA duplex because the mismatch tends to disrupt during the DNA bending, which affects the calculated free energy. Note that in the X-ray DNA-MutS $\zeta$  complex the G/T pair is modestly disrupted, it shows a change in base pair parameter shear, buckle and opening (Fig. 2B) and it is shifted towards minor groove.<sup>30</sup> In our computations, we compared global bending of different DNA systems, with stable mismatch base pair inside the duplex. Since various mismatch pairs have different free energy minima for shear, we set up differently the shear restraint ranges to keep the pair stable. The ranges for each mismatch were chosen based on (i) behaviour of the mismatch in the production 150 ns simulations where we determined the distribution of shear values and (ii) extra ABF simulations, where we probed each mismatch via collective variable shear sampled in the range from  $-5$  to  $5$  Å. Determined shear minima from ABF calculations were in agreement with the distribution of shear values observed in the production simulations (Supplementary Fig. S1). In particular, free energy minima of G/T and A+/C mismatches are  $-2.25$  and  $-2.58$  Å, respectively, so the shear restraint range was set from  $-10$  to  $0$  Å. Shear restraints for other mismatches are indicated in Supplementary Material.

The free energy profiles describing DNA bending were calculated for RMSD in the interval from  $1.45$  to  $5.45$  Å. ABF simulations employed a single window approach, which spans the entire range of the sampled collective variable. This window was split into 41 bins, which serve for numerical evaluation of mean force only. The bins are visited by free diffusion once enough samples are collected in individual bins to obtain a good estimate of mean force, which is subtracting from MM potential force to remove any barriers along the collective variable. Five ABF walkers were commenced from five restart files of the production dynamics each taken after 30 ns to establish the independence of walkers. As result,  $\sim 2$  million samples taken at each MD step were accumulated in each bin, see Supplementary Figure S2. The time step used in our MD simulations was 2 fs. The snapshots were saved to the trajectory file every 5,000 steps (10 ps).

The relaxed DNA is shown as a minimum on the free energy profile (Supplementary Fig. S3), representing a stable thermodynamic state. On the contrary, the bent structure is not thermodynamically stable because of the absence of MutS $\zeta$  in our model. Therefore, we

had to select RMSD value which would represent the bend state and serve to compare coldspots and hotspots. To guess a fair estimate, we have analysed the behaviour of the relaxed DNA state observed in the 150 ns production dynamics. RMSD towards an average DNA structure was calculated over the entire production dynamics. In other words, we have reversed the definition of the target structure for RMSD calculation from the bend conformation to the relaxed conformation. Depending on the system, the RMSD using this target fluctuated from 1 to 4 Å, with maximum occurrence at about  $1.55$  Å for the set of atoms defining used RMSD, see Supplementary Figure S2 in ref.<sup>39</sup> This value was then set as a threshold representing a bend structure to deduct bending free energies from the ABF/MWA simulations.

To evaluate the significance of calculated bending free energies between coldspots and hotspots we calculated errors directly from the ABF/MWA simulations. The error of the mean force [ $\sigma(\text{dG}/\text{dRMSD})$ ] is shown in Supplementary Fig. S4 (left). The error is small and approximately constant in the entire range of RMSD indicating no problem with sampling. The integrated value of errors for the resulting free energy is shown in Supplementary Figure S4 (right). Its value at RMSD threshold of  $1.55$  Å is  $\sim 0.3$  kcal/mol (Supplementary Table S1). We also repeated ABF calculations (i.e. we ran new independent simulations with different starting restart files) for C01, C02, H01 and H02 motifs with G/T. Comparison with original calculations can be found in Supplementary Table S2, the found differences indicate errors in the calculated free bending energies, which is in the same range as provided by the previous analysis, i.e. 0.3 kcal/mol.

The MD trajectories were processed using the ptraj module of AMBER and visualized using the VMD programme.<sup>54</sup> Base pair parameters were calculated with Curves<sup>55</sup> and 3DNA.<sup>56</sup> Base pair geometries in Figure 1 were drawn using the Nemesis programme.<sup>57</sup> Local bend of base pair steps was calculated as described by Sherer et al.<sup>58</sup> based on tilt and roll parameters obtained with the following formula:

$$\text{bend} = \sqrt{\text{roll}^2 + \text{tilt}^2}$$

## 2.4. Logistic regression

We performed logistic regression based on calculated bending free energies of coldspots/hotspots to classify other studied groups of motifs (i.e. motifs with CpG or CpHpH sequence, population hotspots and motifs with *de novo* mutations). The analysis was performed in R. Coldspots were defined as Group 0 and hotspots as Group 1. Obtained *P*-values close to 1 indicate that particular motif behaves as a hotspot while *P*-values close to 0 indicate that motif behaves as coldspot. The classification was performed based on bending free energies for G/T pair, which best-distinguished coldspots and hotspots.

## 2.5. DNA melting curve analysis

We measured a DNA duplex stability for 15-nt long DNA oligonucleotides carrying coldspot/hotspot motifs. We designed the same sequences, which were used in our calculations, i.e. the forward sequence was 5'**GAACC** XXGXX **CTAGG**3', the reverse complementary sequence was (5'**CCTAG** XXTXX **GGTTC**3'). The central region contained coldspot or hotspot with G/T pair (in bold). The mixture of both primers (1  $\mu$ M each; final concentration was tuned to maintain equal starting fluorescence signal for all analysed samples) in 1 mM KCl was denatured (90°C) and slowly (0.1°C/s) cooled to 4°C. Hybridized oligonucleotides were kept on the ice, and EvaGreen dye (Biotium) was added to the final concentration 1 $\times$  (according to the manufacturer's instructions). The melting analysis was performed in the range 20–90°C (0.2°C/s; fluorescence acquired at each temperature step) on Rotor-Gene 6000 (Qiagen). The melting curves were analysed using Melt curve analysis implemented in RotorGene software as the derivative of the raw data.

## 3. Results and discussion

### 3.1. Bending of 10 top coldspots and 10 hotspots with G/T and A+/C mismatches

We used two coldspots (C01-AAGAA and C02-CAGTG) and two hotspots (H01-AGGTA and H02-TGGAA) from our previous study<sup>39</sup> (systems with G/T described in our previous study were recalculated with the new version of PMFLib, which was used in free energy calculations of other systems in this study) and added eight new coldspots and eight new hotspots with lowest Fisher combined *P*-values calculated in our previous study. These *P*-values were determined based on analysis of mutations in five genes (*LDLR*, *PAH*, *CFTR*, *F8* and *F9*) and indicate significance for each motif to be a coldspot or hotspot (see Tables 2 and 3 in<sup>39</sup>). In the case of hotspots, we excluded motifs with CpG dinucleotide in the middle of the 5-nt segment and also CpHpG motifs because mutability of these sequences can be associated with deamination of mC leading to C > T mutations that are primarily targeted by BER pathway (analysis of motifs with CpG and CpHpG was thus conducted separately, see below).

Since coldspots have mostly A-T pair in the middle of the 5-nt segment, whereas hotspots have G = C pair (see sequence logo in Fig. 3 in ref.<sup>39</sup>), different representation of mismatches exists in coldspots and hotspots. In particular, G/T, A+/C, A/G and C/T are equally represented, but A/A and T/T pairs occur more frequently in coldspots while C+/C and G/G exist more often in hotspots. Base pairs G/T and A+/C can substitute the canonical pairs G = C and A-T without large structural perturbation of the duplex.<sup>29,59</sup> Their H-bonding scheme and the C1'–C1' distances are similar to the canonical pairs (Fig. 1). Other non-canonical pairs show larger differences in the

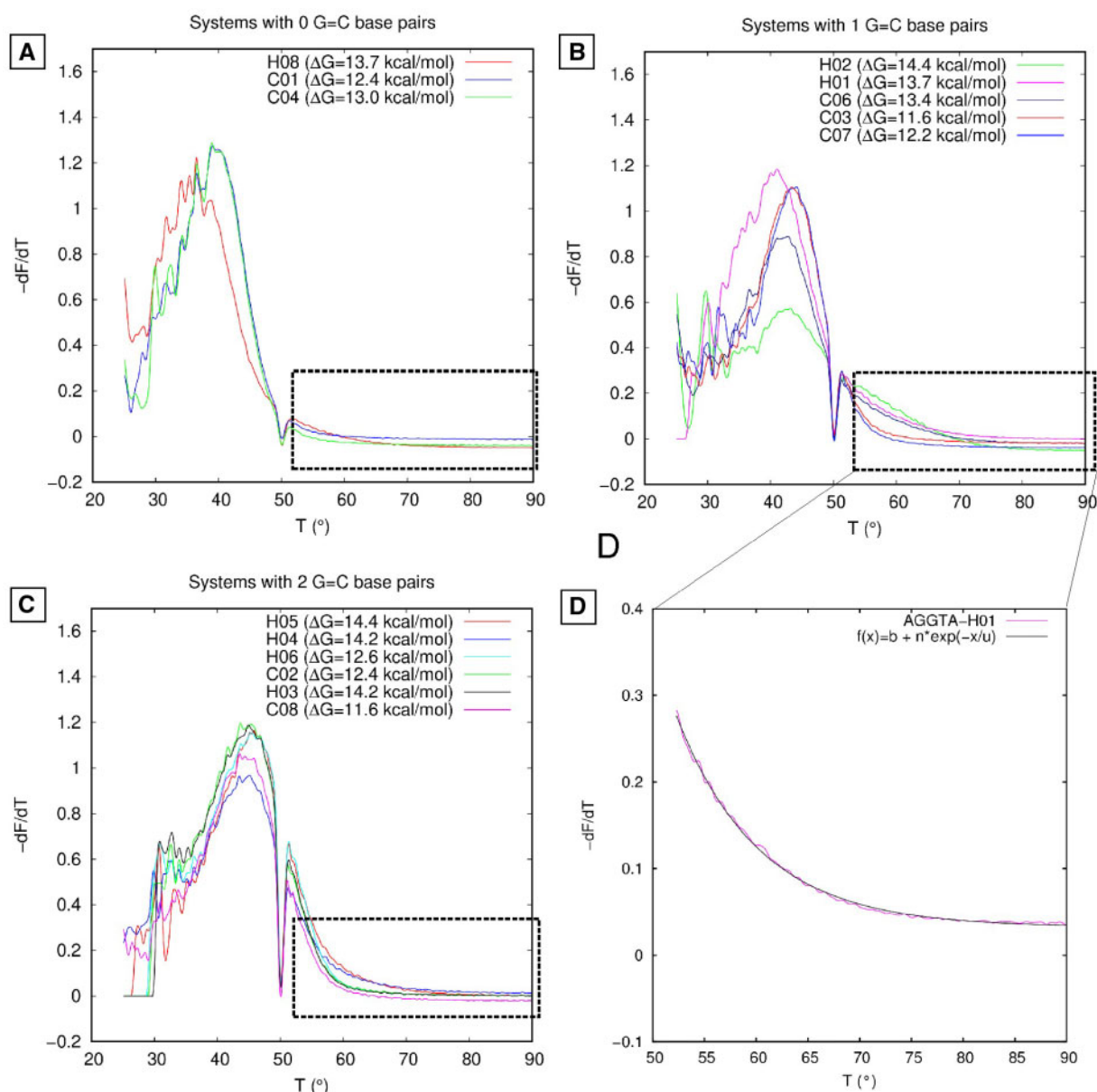
C1'–C1' distances (Fig. 1) and have more variable base pairing.<sup>38</sup> In addition, G/T and A+/C are results of transitions (purine-to-purine or pyrimidine-to-pyrimidine substitutions), which are more frequent than transversions (purine-to-pyrimidine or pyrimidine-to-purine substitutions).<sup>44</sup> Therefore, we primarily focussed on bending analysis of coldspots and hotspots with these mismatches.

Our calculations showed that coldspots with G/T and A+/C pairs are in average more flexible than hotspots about 1.2 and 0.6 kcal/mol, respectively (see Table 1 and Supplementary Fig. S3 where bending free energy profiles are visualized). Considering the error size of calculated free energies, which is around 0.3 kcal/mol (see Supplementary Fig. S4 and Supplementary Tables S1 and S2), the difference between coldspots and hotspots with G/T is significant while the difference between coldspots and hotspots with A+/C is comparable with the error size. We also found several exceptions, e.g. hotspots H06 and H10 show relatively low bending free energies for G/T while some coldspots exhibit higher bending free energies (e.g. C04, C06 and C10, see Table 1). Based on our hypothesis we assume that coldspots are on average (considering all mismatch types) more flexible than hotspots, but we do not expect that every coldspot with any mismatch will be more flexible than any hotspot. Also, we do not consider here that mutability of DNA could be affected by other factors than by sequence elasticity and that MD methods suffer from various approximations. For instance study by LaRosa and Zacharias<sup>60</sup> showed that local untwisting and bending of the DNA lowers the penalty for flipping of modified base (8oxoG), which is recognized by bacterial MutM protein (homologue of human hOGG1) within BER. Most probably, a similar mechanism plays a role in MMR, where DNA bending can be facilitated by disrupted mismatch base pair. However, in this study, we restrained the central mismatch base pair inside duplex because it opened spontaneously only in some ABF simulations, which prevented comparison of the global bending properties among different systems. Thus, our analysis aimed to detect a potential difference in global elasticity between coldspots and hotspots.

Based on logistic regression, we calculated *P*-values for hotspots and coldspots (Table 1). It can be seen that *P*-values of the most hotspots are around 0.8 or 0.9, only H06, H07 and H10 motifs exhibit lower values (Table 1). In the case of coldspots, six motifs exhibit *P*-values below 0.3, other four motifs: C04, C06, C09 and C10 have *P*-values between 0.46 and 0.68, so none of them behaves as a typical hotspot.

### 3.2. Bending of 10 top coldspots and 10 hotspots with other mismatches

In the bending analysis of G/A, C/T, A/A, C+/C, T/T and G/G, we encountered a problem with the stability of some mismatched base pairs. We observed disruptions of symmetrical A/A, G/G and C+/C mismatches during bending which lasted for several ns during the ABF simulations. Although shear of the mismatch was restrained (as described in Supplementary Material), the mismatch was perturbed in such extent that its bases stack on top of each other (Supplementary Fig. S5). Our tests revealed that DNA with a disrupted mismatch has lower bending free energy than when the pair is perfectly stable (observed for DNA with G/T pair). This indicates that DNA with perturbed pair can be bent more easily, which correlates with observation in ref.<sup>35</sup> where abasic DNA compared with DNA with various mismatches exhibited lowest bending free energy. We were not able to stabilize these mismatches in the ABF simulations. Therefore, we did not derive bending free energies for these



**Figure 3.** Melting curves of oligonucleotide duplexes with the G/T mismatch acquired using RotorGene software, the derivative of the raw fluorescence with respect to temperature ( $dF/dT$ ) is shown. (A) systems with 0 G = C base pair, (B) systems with 1 G = C base pair and (C) systems with 2 G = C base pairs. Bending free energies of analysed motifs are indicated. (D) The melting curve fitted by exponential function in the range of 52–90°C.

systems. Other mismatches C/T, A/G and T/T were stable during bending. In the case of C/T and A/G mismatches, we did not detect a significant difference between bending free energies of coldspots and hotspots, although coldspots were slightly more flexible than hotspots (Supplementary Table S3). In the case of T/T pair we observed the opposite trend, i.e. hotspots were more flexible than coldspots (Supplementary Table S3). However, it should be noted that this mismatch is not equally distributed in coldspots and hotspots, thus we analysed only four hotspots with T/T and eight coldspots with T/T. In the available X-ray complexes of DNA/MutS A/A, G/G and A/C pairs exist in *anti-syn* conformation,<sup>61,62</sup> only G/T pair was detected in *anti-anti* (Fig. 2).<sup>30</sup> Most probably the initial *anti-anti* form, which is prevailing in the free form of DNA, is not suitable geometry for DNA complexed with protein.

Although C/T, A/G and T/T pairs stayed stable in the ABF simulations, we assume that mismatches less compatible with the canonical pairs require base-pair rearrangement during bending to optimize its geometry. Keeping mismatch in one base pair geometry can create tension which might affect bending free energy. It is also possible that rearrangement of the A/C mismatch during bending (we performed bending analysis with A+/C in *anti-anti* form, while there is unprotonated *anti-syn* form in the X-ray structure of DNA/MutS) would increase a difference in bending free energies between coldspots and hotspots. However, description of controlled mismatch rearrangement during bending would require more collective variables, which would be very computationally demanding. In the next sections, we analysed the bending of other DNA motifs but only with G/T and A+/C pairs.

**Table 1.** Bending free energies of top 10 coldspots and top 10 hotspots with G/T and A+/C pairs

ID	Motif 5'→3'/5'→3'	Fisher combined <i>P</i> -value <sup>39</sup>	Bending free energy for G/T (kcal/mol)	Bending free energy for A+/C (kcal/mol)	<i>P</i> -value from logistic regression
Coldspots					
C01	AAGAA/TTCTT	1.27E-06	12.4	12.4	0.17
C02	CAGTG/CACTG	9.22E-05	12.4	12.8	0.17
C03	AAAGA/TCTTT	9.37E-09	11.6	12.3	0.03
C04	AAAAT/ATTTT	1.08E-08	13.0	14.1	0.46
C05	AAAAA/TTTTT	1.36E-08	11.9	12.7	0.06
C06	AGAAA/TTTCT	1.78E-08	13.4	13.9	0.68
C07	GAAAA/TTTTC	4.50E-08	12.2	13.1	0.12
C08	GAAGA/TCTTC	6.47E-07	11.6	12.5	0.03
C09	GGAGA/TCTCC	1.22E-06	13.2	12.7	0.59
C10	GGAAA/TTTCC	3.04E-06	13.2	14.0	0.55
Average values			12.5	13.1	
Hotspots					
H01	AGGTA/TACCT	1.24E-11	13.7	14.0	0.80
H02	TGGAA/TTCCA	6.58E-04	14.4	14.5	0.95
H03	AGGTG/CACCT	2.44E-03	14.2	13.6	0.93
H04	TGAGT/ACTCT	1.37E-03	14.2	14.5	0.93
H05	TGGCT/AGCCA	8.02E-04	14.4	14.7	0.96
H06	ACATG/CATGT	5.38E-04	12.6	13.2	0.26
H07	GGGCA/TGCCC	4.70E-03	13.0	13.3	0.48
H08	TTGTA/TACAA	4.79E-03	13.7	12.9	0.82
H09	GGATG/CATCC	4.90E-03	13.6	13.3	0.78
H10	GCATG/CATGC	7.46E-03	12.5	12.5	0.22
Average values			13.6	13.7	

Bending free energies smaller than 13.0 kcal/mol are in grey fields.

**Table 2.** Bending free energies of hotspots with CpG and CpHpG sequence

Type	Motif 5'→3'/5'→3'	Fisher combined <i>P</i> -value <sup>39</sup>	Bending free energy for G/T (kcal/mol)	Bending free energy for A+/C (kcal/mol)	<i>P</i> -value from logistic regression
CpG	ACGGC/GCCGT	5.51E-05	12.0	12.4	0.08
CpG	CCGAG/CTCGG	2.11E-05	13.1	11.8	0.52
CpG	TCGCA/TGCGA	2.69E-08	13.2	13.1	0.58
CpHpG	CTGTG/CACAG	2.97E-06	13.0	11.8	0.46
CpHpG	CTGGG/CCCAG	2.24E-05	12.8	11.5	0.35
CpHpG	CAGTA/TACTG	2.87E-04	12.3	12.8	0.14
Average values			12.7	12.2	

Bending free energies smaller than 13.0 kcal/mol are in grey fields.

### 3.3. Bending of CpG and CpHpG motifs with G/T and A+/C

In our previous study, motifs with CpG or CpHpG were frequently detected as hotspots; particularly, these motifs represented 14 out of top 20 hotspots (see Table 3 in ref.39). Thus, we separately analysed three hotspots with CpG and three hotspots with CpHpG possessing G/T and A+/C pairs (Table 2). These hotspots have low Fisher combined *p*-values<sup>39</sup> and belong among top 10 hotspots in our previous study.<sup>39</sup> Calculated bending free energies of these systems are low and correspond to coldspots rather than hotspots, particularly, average bending free energies were 12.7 and 12.2 kcal/mol for G/T and A+/C, respectively. Calculated *P*-values from the logistic regression were below or around 0.5. This correlates with previous computational studies<sup>31,63–65</sup> where pyrimidine–purine and purine–purine (or

pyrimidine–pyrimidine) dinucleotides were found to be more flexible than purine–pyrimidine.

### 3.4. Bending of DNA with Czech population hotspots

We investigated the bending properties of 10 population hotspots detected in Czech patients. Theoretically, a DNA site can become a population hotspot due to only one mutation event, when the arisen mutation is spread throughout the population. Thus, bending properties of these sequences do not have to exhibit stiffness which we detected in hotspots described in Table 1.

Our cooperating genetic laboratory (in Centre of Molecular Biology and Gene Therapy, University Hospital Brno) analyses mainly genes associated with neuromuscular, skin and metabolic diseases. Therefore, we focussed on the most significant Czech

**Table 3.** Bending free energies of 10 Czech population hotspots with G/T and A+/C

Gene	Mutation at protein, DNA level, transcript number	Number of alleles <sup>a</sup>	Motif 5'→3'/5'→3'	Bending free energy for G/T (kcal/mol)	Bending free energy for A+/C (kcal/mol)	P-value from logistic regression
<i>ALOX12b</i>	p.Tyr521cys, c.1562A>G, NM_001139.2	6/12	TTACC/GGTAA	14.0	14.1	0.89
<i>ALOXE3</i>	p.Arg234Term, c.700C>T, NM_021628.2	6	TTCGA/TCGAA	13.1	14.4	0.51
<i>TGM1</i>	p.Arg142His, c.425G>A, NM_000359.2	1/2	GCGCC/GGCGC	12.5	13.7	0.21
<i>CAPN3</i>	p.Arg448His, c.1343G>A, NM_000070.2	2/4	CCGCA/TGCGG	13.0	12.4	0.43
<i>ANO5</i>	p.Arg758Cys, c.2272C>T, NM_213599.2	1/5	CCCGT/ACGGG	11.8	12.6	0.05
<i>SGCA</i>	p.Arg77Cys, c.229C>T, NM_000023.3	4/6	TCCGC/GCGGA	11.7	11.8	0.04
<i>LDLR</i>	p.Gly592Glu, c.1775G>A, NM_000527.4	104/188	CGGGG/CCCCG	12.4	12.7	0.17
<i>PAH</i>	p.Arg408Trp, c.1222C>T, NM_000277.2	560	CTCGG/CCGAG <sup>b</sup>	13.1	11.8	0.52
<i>ATP7B</i>	p.Trp779Term, c.2336G>A, NM_000053.3	14	GTGGC/GCCAC	13.1	12.0	0.49
<i>CLCN1</i>	p.Arg894Term, c.2680C>T, NM_000083.2	38	CTCGA/TCGAG	13.1	13.9	0.52
Average values				12.8	12.9	

CpG or CpHpG motifs are in bold and energy below 13 kcal/mol is in the grey field.

<sup>a</sup>Number of alleles indicated in our publications (see above)/number of alleles detected up to date if differ.

<sup>b</sup>This motif is already included among top hotspots with CpG see Table 2.

population hotspots in these genes detected in our previous studies. In particular, we analysed hotspots within 5-nt segments in: *TGM1*, *ALOXE3*, *ALOX12b* genes associated with congenital ichthyoses,<sup>40</sup> in *CAPN3*, *SGCA* and *ANO5* genes associated with limb-girdle muscular dystrophies,<sup>66</sup> in *LDLR* gene associated with familial hypercholesterolaemia,<sup>67</sup> in *PAH* gene associated with phenylketonuria,<sup>68</sup> in *ATP7B* gene associated with Wilson disease<sup>69</sup> and in *CLCN1* gene associated with Myotonia Congenita<sup>70</sup> (Table 3). As mentioned earlier we analysed only population hotspots leading to transitions. Many analysed population hotspots contain CpG or CpHpG (8 out of 10). It can be seen that five motifs with CpG or CpHpG have low bending free energies (around or below 13 kcal/mol) for both mismatches, only three motifs with CpG (in *ALOXE3*, *TGM1* and *CLCN1*) exhibited higher bending free energies for one mismatch (Table 3). In the case of motif in *ALOX12b*, which does not contain CpG or CpHpG sequence, its bending free energies are high for both mismatches (14.0 kcal/mol for G/T and 14.1 kcal/mol for A+/C), the second non-CpG motif (in *ATP7B* gene); however, has low bending free energies for both mismatches. Thus, we can see a quite wide range of bending free energies, but on average, bending free energies of these systems are low. P-values derived from the logistic regression are similar to CpG/CpHpG motifs, i.e. 9 out of 10 motifs have a P-value below 0.53 (Table 3), which indicates that bending properties of these motifs are closer to coldspots described in Table 1.

### 3.5. Bending of DNA with *de novo* mutations

Analysis of coldspots/hotspots with G/T and A+/C showed that mutations arise more often in the rigid sequences. To test this observation, we performed bending analysis of 15 *de novo* mutations within 5-nt segments on non-CpG/non-CpHpG sites and we selected only mutations leading to G/T and A+/C pairs. Mutations from *ADA*, *IL2RG*, *SERPING1*, *STAT3* and *WAS* were provided by the cooperating department at the Centre for Cardiovascular Surgery and Transplantation, Brno, particularly, mutations in *ADA*, *SERPING1* and *STAT3* were entirely novel (not published before), mutations in *IL2RG* and *WAS* were recently published by the cooperating department.<sup>71</sup> *ADA*, *IL2RG*, *STAT3* and *WAS* genes are associated with immunodeficiencies, whereas *SERPING1* is associated

with hereditary angioedema. Further, eight novel mutations were taken from the *PAH* gene. In our previous study,<sup>68</sup> we described six novel missense mutations. Three of them (transitions located on non-CpG sequences: p.Asp229Gly, p.Phe263Ser and Ile406Met) were included among the 15 *de novo* mutations (Table 4). Other five mutations in *PAH* were taken from the HGMD database from the beginning of the protein sequence (p.Met1Ile, p.Ser16Pro, p.Ser40Leu, p.Lys42Glu and p.Val45Ala). In the study,<sup>70</sup> we reported two novel mutations in *CLCN1* gene leading to transitions: missense mutation c.905A>G, p.(Tyr302Cys) and splicing mutation c.433 + 3A>G, the latter is located on the GTAGG motif, which represents known indel hotspot derived from GTAAGT.<sup>18</sup> Since GTAGG motif is associated with double base pair substitutions (22%), we analysed only the p.(Tyr302Cys) mutation. The last mutation was taken from *CYP4F22* gene where we recently reported two missense mutations (c.844C>T, p.R282W and c.1085G>A, p.R362Q).<sup>40</sup> However, because both of them are positioned on the CpG dinucleotide, we selected the most recent reported mutation (c.667C>T, p.Q223Term) in this gene in HGMD.<sup>72</sup>

Calculations showed that bending free energies of these motifs are somewhat higher with average value 13.6 kcal/mol for G/T and 13.4 kcal/mol for A+/C (Table 4) which is comparable to hotspots in Table 1. Low bending free energies for both mismatches can be seen only for p.His115Arg in *WAS* gene, p.Ser16Pro in *PAH* gene and p.Tyr302Cys in *CLCN1* gene. Other motifs have bending free energies for at least one mismatch around or above average hotspot value. Logistic regression revealed that except for the three mentioned mutations (p.His115Arg, p.Ser16Pro and p.Tyr302Cys) P-values are higher than 0.52 (seven motifs have P-value even higher than 0.8) (Table 4).

Although we do not have large statistical datasets for studied groups, we can see apparent trends in bending free energies calculated for G/T and A+/C pairs. In the case of G/T pair, average bending free energies of coldspots, CpG/CpHpG motifs and motifs with population hotspots are similar and lower compared with the bending free energies of hotspots and motifs with *de novo* mutations. This trend is less evident for A+/C mismatch where CpG/CpHpG motifs show lowest average bending free energies, and where a smaller difference between coldspots and hotspots exists (this



**Table 4.** Bending free energies of motifs with 15 *de novo* mutations calculated for G/T and A+/C mismatches

Gene	Mutation at protein level, DNA level and transcript number	Motif 5'→3'/5'→3'	Bending free energy for G/T (kcal/mol)	Bending free energy for A+/C (kcal/mol)	P-value from logistic regression
ADA	p.Val130Met, c.388G>A, NM_000022	TGGTG/CACCA	14.2	14.2	0.93
IL2RG	p.Gly232Glu, c.695G>A, NM_000206.2	TGGAA/TTCCA <sup>a</sup>	14.4	14.5	0.95
SERPING1	p.Leu349Pro, c.1046T>C, NM_000062.2	GCTCT/AGAGC	13.2	13.5	0.56
STAT3	p.Lys658Glu, c.1972A>G, NM_139276.2	ATAAG/CTTAT	14.8	13.3	0.98
WAS	His115Arg, c.344A>G, NM_000377	CCACA/TGTGG	12.4	13.1	0.18
PAH	p.Asp229Gly, c.686A>G, NM_000277.2	AGACG/CGTCT	13.3	13.4	0.62
PAH	p.Phe263Ser, c.788T>C, NM_000277.2	CTTCC/GGAAG	13.1	13.8	0.51
PAH	Ile406Met, c.1218A>G, NM_000277.2	ATACC/GGTAT	13.8	13.3	0.83
PAH	p.Met11le, c.3G>A, NM_000277.2	ATGTC/GACAT	14.1	13.6	0.91
PAH	p.Ser16Pro, c.46T>C, NM_000277.2	TCTCT/AGAGA	12.9	12.9	0.38
PAH	p.Ser40Leu, c.119C>T, NM_000277.2	CTCAC/GTGG	13.9	12.7	0.86
PAH	p.Lys42Glu, c.124A>G, NM_000277.2	TCAAA/TTTGA	14.1	13.1	0.92
PAH	p.Val45Ala, c.134T>C, NM_000277.2	AGTTG/CAACT	13.2	14.2	0.58
CLCN1	p.Tyr302Cys, c.905A>G, NM_000083.2	CTACT/AGTAG	13.1	12.9	0.51
CYP4F22	p.Gln223Term, c.667C>T, NM_173483.3	GCCAA/TTGGC	13.6	12.8	0.78
Average values			13.6	13.4	

Bending free energies smaller than 13.0 kcal/mol are in grey fields.

<sup>a</sup>This motif is already included in Table 1 as H02.

correlates with the fact that bending free energies for G/T were taken as the best classifier in logistic regression).

### 3.6. Melting experiments for selected coldspots and hotspots with G/T

Previous experimental study showed that A-RNA has a higher melting temperature ( $T_m$ ) than B-DNA,<sup>73</sup> which correlates with computational studies<sup>74,75</sup> where A-RNA was found to be more stable than B-DNA. Moreover, B-DNA exhibited greater deformability of its backbone than A-RNA, and it was more easily deformed in the global parameters roll and tilt, which were used for calculation of global bending. We measured and analysed melting curves of coldspots and hotspots using thermal denaturation experiments to determine if there is a correlation with calculated bending free energies.

First eight coldspots (C01–C08) and first eight hotspots (H01–H08) were analysed with the central G/T mismatch. Since C05-AAAAA after introducing G/T pair changes sequence to AAGAA, which is identical to C01-AAGAA, we analysed only C01 motif in this experiment. As expected, positions of peaks and derived  $T_m$  were affected by the number of G = C pairs included in the analysed motifs (Supplementary Fig. S6). Therefore, we separated melting curves according to the number of G = C pairs and compared coldspots and hotspots with 0 G = C, 1 G = C and 2 G = C pairs. For system H07 (GGGCA), which contains 3 G = C pairs, we did not have coldspot to compare it. Therefore this system was not considered.

In the plots, we observed a trend where derivative melting curves of rigid motifs decrease slower compare to systems with increased flexibility. This can be seen in the range of 50–70°C. Although this trend is not 100% accurate, it is seen for many motifs in 0 G = C, 1 G = C and 2 G = C groups. For instance, in the case of motifs with 0 G = C (Fig. 3A), H08 motif with highest bending free energy (13.7 kcal/mol) decreases slower than other two systems C01 and C04 with bending free energies 12.4 and 13.0 kcal/mol, respectively. Considering systems with 1 G = C pair (Fig. 3B), the melting curves of motifs decrease almost perfectly according to decreasing bending

free energy. The order of decrease is following: H02 with highest bending free energy (14.4 kcal/mol), H01 (bending free energy 13.7 kcal/mol), C06 (bending free energy 13.2 kcal/mol), C03 (bending free energy 11.6 kcal/mol) and C07 (bending free energy 12.2 kcal/mol). For motifs with 2 G = C pairs (Fig. 3C), we also observed the trend except for H03, which decreased slowly as other coldspots although its bending is 14.2 kcal/mol.

Further, to compare melting curves quantitatively, each one was fitted by the exponential function in the range of 52–90°C (Fig. 3D). Calculated coefficients  $b$ ,  $n$  and  $u$  were derived. The coefficient  $b$  represents intercept-like coefficient, and its value is around 0 for all systems while coefficients  $n$  and  $u$  describe the curvature of the melting curves. Relative comparison of  $n$  and  $u$  within the groups (0 G = C, 1 G = C and 2 G = C) shows that systems with slow decrease have smaller  $n$  but higher  $u$  (Supplementary Table S4).

### 3.7. Analysis of local bend, twist and three translational base pair step parameters for 10 coldspots and 10 hotspots with G/T in unbiased 150 ns simulations

Analysis of local bend revealed no significant differences in the average local bend on central Steps 7 and 8 (containing mismatch base pair) between coldspots and hotspots. However, local bend values of the central Step 7 correlated with the calculated bending free energies. In particular, coldspots C04, C06, C09 and C10 with higher bending free energies (13.0, 13.4, 13.2 and 13.2 kcal/mol) showed lower local bend on this step (8.2, 7.9, 7.8 and 7.7°, respectively) than other coldspots (Supplementary Table S5). Similarly, hotspots H06 and H10 with low bending free energies (12.6 and 12.5 kcal/mol) exhibited higher local bend on this step (10 and 10.8°, respectively) than other hotspots which are stiffer (Supplementary Table S6). Exception is H08, of which bending energy is 13.7 kcal/mol and has also larger local bend (10.7°).

In the case of twist, central Steps 7 and 8 in coldspots with average values 25.8 and 41.1° deviate more from standard twist value (ca. 36°) than corresponding steps in hotspots where central Steps 7

and 8 exhibit values 28.9° and 37.8° (Supplementary Tables S7 and S8). Analysis of three translational base pair step parameters (shift, slide and rise) revealed no significant differences between coldspots and hotspots and no correlation with the bending free energies (Supplementary Tables S9–S14).

It should be noted that biased ABF simulations exhibit two different modes of bending, see graphs of mean force in Figure S4 left. In particular, there are two linear regions, each representing a harmonic mode of bending with different force constants. One is observed in the range from 5.45 to ca 2.5 Å and the second more profound is in the range from 2.5 to 1.5 Å. The bending in the second region has major impact on discrimination between coldspots and hotspots. Basically, the unbiased simulations can reveal information only about the first mode. The second mode, which seems to be more important, is inaccessible in unbiased simulations and it can be expected that important features cannot be deduced from analysis of such MD trajectories.

#### 4. Conclusion

We analysed the bending properties of damaged DNA in different sequence contexts employing biased MDs simulations. We tested sequences, which are frequently associated with germline mutations (hotspots) and rarely associated with germline mutations (coldspots). Main attention was paid to motifs with G/T, and A+/C mismatches as these pairs are the result of transitions and have base pair geometry well defined in the duplex structure. Despite the limitation of the employed model, our results revealed the relation between mutability of DNA sequences and their bending properties, particularly, coldspots were found to be more flexible than hotspots (Table 1). This is in agreement with the suggested model of the MMR process,<sup>22</sup> in which more flexible segments are more effectively repaired.

Our attempt to analyse other mismatch types including A/A, G/G and C+/C failed because biased simulations were not stable and thus derived free energy profiles were affected by other than bending movements. In the case of C/T, A/G and T/T mismatches, the trend in bending free energies between coldspots and hotspots was not convincing. Coldspots with C/T and A/G were found to be slightly more flexible than hotspots, while the opposite trend was detected for T/T mismatch (Supplementary Table S3). This might be solved by employing more collective variables, which would also describe mismatch rearrangement during bending. However, it can be expected that such simulations would be very computationally demanding. Apart from DNA elasticity, detection of mismatches and subsequent repair process in the cell is most likely influenced by other factors, which we did not consider in our model. Thus, we cannot expect a perfect correlation between DNA mutability and elasticity. Moreover, mismatch recognition is fine-tuned by interaction with amino acids phenylalanine 432 and glutamine 434 from MutS protein.<sup>30</sup> Our model explaining the emergence of mutations in DNA is therefore simplistic, but it suggests that DNA elasticity plays an important role.

Bending analysis of motifs with CpG and CpHpG revealed that bending free energies of these systems are low (Table 2), which indicates that their mutability is more likely related to spontaneous deamination of mC leading to C > T mutations, which are targeted by BER pathway. Similarly, the bending free energies of the Czech population hotspots were lower (Table 3). These motifs often contained CpG and CpHpG sequences. In contrast, motifs with *de novo* mutations (intentionally selected non-CpG and non-CpHpG sites)

exhibited higher bending free energies (Table 4) that were comparable with bending free energies of hotspots (Table 1). This observation supports our hypothesis, which suggests that mutations appear more likely on the rigid DNA sites.

Our computational results are further complemented by thermal denaturation experiments where we measured the stability of selected coldspots and hotspots with the G/T mismatch. Obtained results showed that melting curves of hotspots decreased slower than melting curves of coldspots in the range of 50–70°C, so the hotspots exhibit higher stability that could be associated with increased rigidity detected in our calculations.

#### Acknowledgements

We would like to thank Hana Grombřířková from the Centre for Cardiovascular Surgery and Transplantation, Brno for providing novel mutations in *ADA*, *IL2RG*, *SERPING1*, *STAT3* and *WAS* genes.

#### Supplementary data

Supplementary data are available at DNARES online.

#### Funding

Computational resources were provided by the CESNET (LM2015042) and the CERIT Scientific Cloud (LM2015085), provided under the programme 'Projects of Large Research, Development and Innovations Infrastructures'. This work was supported by the Ministry of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project 'IT4Innovations National Supercomputing Center (LM2015070)'. This research was financially supported by the Grant Agency of the Czech Republic (GA16-11619S/2016) and by the Ministry of Education, Youth and Sports of the Czech Republic under the project CEITEC 2020 (LQ1601).

#### Conflict of interest

None declared.

#### References

- Roach, J.C., Glusman, G., Smit, A.F.A., et al. 2010, Analysis of genetic inheritance in a family quartet by whole-genome sequencing, *Science*, **328**, 636–9.
- Lynch, M. 2010, Rate, molecular spectrum, and consequences of human mutation, *Proc. Natl. Acad. Sci. USA*, **107**, 961–8.
- Nabel, C.S., Manning, S.A. and Kohli, R.M. 2012, The curious chemical biology of cytosine: deamination, methylation, and oxidation as modulators of genomic potential, *ACS Chem. Biol.*, **7**, 20–30.
- Lewis, C.A., Crayle, J., Zhou, S.T., Swanstrom, R. and Wolfenden, R. 2016, Cytosine deamination and the precipitous decline of spontaneous mutation during Earth's history, *Proc. Natl. Acad. Sci. USA*, **113**, 8194–9.
- Koag, M.C., Nam, K. and Lee, S. 2014, The spontaneous replication error and the mismatch discrimination mechanisms of human DNA polymerase beta, *Nucleic Acids Res.*, **42**, 11233–45.
- Iyer, R.R., Pluciennik, A., Burdett, V. and Modrich, P.L. 2006, DNA mismatch repair: functions and mechanisms, *Chem. Rev.*, **106**, 302–23.
- Rogozin, I.B. and Pavlov, Y.I. 2003, Theoretical analysis of mutation hotspots and their DNA sequence context specificity, *Mut. Res.*, **544**, 65–85.
- Cooper, D.N., Bacolla, A., Ferec, C., Vasquez, K.M., Kehrer-Sawatzki, H. and Chen, J.M. 2011, On the sequence-directed nature of human gene mutation: the role of genomic architecture and the local DNA sequence

- environment in mediating gene mutations underlying human inherited disease, *Hum. Mutat.*, **32**, 1075–99.
9. Youssoufian, H., Kazazian, H.H., Phillips, D.G., Aronis, S., Tsiftis, G., Brown, V.A. and Antonarakis, S.E. 1986, Recurrent mutations in hemophilia-a give evidence for CpG mutation hotspots, *Nature*, **324**, 380–2.
  10. Walsh, C.P. and Xu, G.L. 2006, Cytosine methylation and DNA repair, *Curr. Top. Microbiol. Immunol.*, **301**, 283–315.
  11. Jacobs, A.L. and Schar, P. 2012, DNA glycosylases: in DNA repair and beyond, *Chromosoma*, **121**, 1–20.
  12. Cooper, D.N. and Krawczak, M. 1993, *Human gene mutation*. BIOS Scientific Publishers: Oxford.
  13. Neuberger, M.S., Harris, R.S., Di Noia, J. and Petersen-Mahrt, S.K. 2003, Immunity through DNA deamination, *Trends Biochem. Sci.*, **28**, 305–12.
  14. Ollila, J., Lappalainen, I. and Vihinen, M. 1996, Sequence specificity in CpG mutation hotspots, *FEBS Lett.*, **396**, 119–22.
  15. Bird, A.P. 1980, DNA methylation and the frequency of CpG in animal DNA, *Nucleic Acids Res.*, **8**, 1499–504.
  16. Hodgkinson, A. and Eyre-Walker, A. 2011, Variation in the mutation rate across mammalian genomes, *Nat. Rev. Genet.*, **12**, 756–66.
  17. Cooper, D.N., Mort, M., Stenson, P.D., Ball, E.V. and Chuzhanova, N.A. 2010, Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides, *Hum. Genomics*, **4**, 406–10.
  18. Chuzhanova, N.A., Anassis, E.J., Ball, E.V., Krawczak, M. and Cooper, D.N. 2003, Meta-analysis of indels causing human genetic disease: mechanisms of mutagenesis and the role of local DNA sequence complexity, *Hum. Mutat.*, **21**, 28–44.
  19. Modrich, P. 1991, Mechanisms and biological effects of mismatch repair, *Annu. Rev. Genet.*, **25**, 229–53.
  20. Cho, M., Han, M.S. and Ban, C. 2008, Detection of mismatched DNAs via the binding affinity of MutS using a gold nanoparticle-based competitive colorimetric method, *Chem. Commun.*, 4573–5.
  21. Brown, J., Brown, T. and Fox, K.R. 2001, Affinity of mismatch-binding protein MutS for heteroduplexes containing different mismatches, *Biochem. J.*, **354**, 627–33.
  22. Gorman, J., Chowdhury, A., Surtees, J.A., Shimada, J., Reichman, D.R., Alani, E. and Greene, E.C. 2007, Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6, *Mol. Cell*, **28**, 359–70.
  23. Qiu, R., DeRocco, V.C., Harris, C., Sharma, A., Hingorani, M.M., Erie, D.A. and Wenginger, K.R. 2012, Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling, *EMBO J.*, **31**, 2528–40.
  24. Jeong, C., Cho, W.-K., Song, K.-M., et al. 2011, MutS switches between two fundamentally distinct clamps during mismatch repair, *Nat. Struct. Mol. Biol.*, **18**, 379–U174.
  25. Groothuizen, F.S., Winkler, I., Cristovao, M., et al. 2015, MutS/MutL crystal structure reveals that the MutS sliding clamp loads MutL onto DNA, *Elife*, **4**, e06744.
  26. Hingorani, M.M. 2016, Mismatch binding, ADP-ATP exchange and intramolecular signaling during mismatch repair, *DNA Repair (Amst.)*, **38**, 24–31.
  27. Lamers, M.H., Perrakis, A., Enzlin, J.H., Winterwerp, H.H.K., de Wind, N. and Sixma, T.K. 2000, The crystal structure of DNA mismatch repair protein MutS binding to a G center dot T mismatch, *Nature*, **407**, 711–7.
  28. Obmolova, G., Ban, C., Hsieh, P. and Yang, W. 2000, Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA, *Nature*, **407**, 703–10.
  29. Wang, H., Yang, Y., Schofield, M.J., et al. 2003, DNA bending and unbending by MutS govern mismatch recognition and specificity, *Proc. Natl. Acad. Sci. USA*, **100**, 14822–7.
  30. Warren, J.J., Pohlhaus, T.J., Changela, A., Iyer, R.R., Modrich, P.L. and Beese, L.S. 2007, Structure of the human MutS alpha DNA lesion recognition complex, *Mol. Cell*, **26**, 579–92.
  31. Perez, A., Lankas, F., Luque, F.J. and Orozco, M. 2008, Towards a molecular dynamics consensus view of B-DNA flexibility, *Nucleic Acids Res.*, **36**, 2379–94.
  32. Kriegel, F., Matek, C., Drsata, T., et al. 2018, The temperature dependence of the helical twist of DNA, *Nucleic Acids Res.*, **46**, 7998–8009.
  33. Drsata, T., Spackova, N., Jurecka, P., Zgarbova, M., Sponer, J. and Lankas, F. 2014, Mechanical properties of symmetric and asymmetric DNA A-tracts: implications for looping and nucleosome positioning, *Nucleic Acids Res.*, **42**, 7383–94.
  34. Hospital, A., Faustino, I., Collepardo-Guevara, R., Gonzalez, C., Gelpi, J.L. and Orozco, M. 2013, NAFlex: a web server for the study of nucleic acid flexibility, *Nucleic Acids Res.*, **41**, W47–W55.
  35. Sharma, M., Predeus, A.V., Mukherjee, S. and Feig, M. 2013, DNA bending propensity in the presence of base mismatches: implications for DNA repair, *J. Phys. Chem. B*, **117**, 6194–205.
  36. Curuksu, J., Zacharias, M., Lavery, R. and Zakrzewska, K. 2009, Local and global effects of strong DNA bending induced during molecular dynamics simulations, *Nucleic Acids Res.*, **37**, 3766–73.
  37. Curuksu, J., Zakrzewska, K. and Zacharias, M. 2008, Magnitude and direction of DNA bending induced by screw-axis orientation: influence of sequence, mismatches and abasic sites, *Nucleic Acids Res.*, **36**, 2268–83.
  38. Rossetti, G., Dans, P.D., Gomez-Pinto, I., Ivani, I., Gonzalez, C. and Orozco, M. 2015, The structural impact of DNA mismatches, *Nucleic Acids Res.*, **43**, 4309–21.
  39. Ruzicka, M., Kulhanek, P., Radova, L., et al. 2017, DNA mutation motifs in the genes associated with inherited diseases, *PLoS One*, **12**, e0182377.
  40. Buckova, H., Noskova, H., Borska, R., et al. 2016, Autosomal recessive congenital ichthyoses in the Czech Republic, *Br. J. Dermatol.*, **174**, 405–7.
  41. Chien, Y.H., Chiang, S.C., Huang, A., et al. 2004, Mutation spectrum in Taiwanese patients with phenylalanine hydroxylase deficiency and a founder effect for the R241C mutation, *Hum. Mutat.*, **23**, 206.
  42. Chiu, Y.H., Chang, Y.C., Chang, Y.H., et al. 2012, Mutation spectrum of and founder effects affecting the PTS gene in East Asian populations, *J. Hum. Genet.*, **57**, 145–52.
  43. Darve, E., Rodríguez-Gómez, D. and Pohorille, A. 2008, Adaptive biasing force method for scalar and vector free energy calculations, *J. Chem. Phys.*, **128**, 144120.
  44. Keller, I., Bensasson, D. and Nichols, R.A. 2007, Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes, *PLoS Genet.*, **3**, 185–91.
  45. Case, D.A., Babin, V., Berryman, J.T., et al. 2014, *AMBER 14*, University of California: San Francisco.
  46. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. and Klein, M. 1983, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, **79**, 926–35.
  47. Joung, I.S. and Cheatham, T.E. 2008, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations, *J. Phys. Chem. B*, **112**, 9020–41.
  48. Ivani, I., Dans, P.D., Noy, A., et al. 2016, Parmbsc1: a refined force field for DNA simulations, *Nat. Methods*, **13**, 55–8.
  49. Zhou, H.Q., Hintze, B.J., Kimsey, I.J., et al. 2015, New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey, *Nucleic Acids Res.*, **43**, 3420–33.
  50. Johnson, S.J. and Beese, L.S. 2004, Structures of mismatch replication errors observed in a DNA polymerase, *Cell*, **116**, 803–16.
  51. Skelly, J.V., Edwards, K.J., Jenkins, T.C. and Neidle, S. 1993, Crystal-structure of an oligonucleotide duplex containing G.G base-pairs - influence of mispairing on DNA backbone conformation, *Proc. Natl. Acad. Sci. USA*, **90**, 804–8.
  52. Raiteri, P., Laio, A., Gervasio, F., Micheletti, C. and Parrinello, M. 2006, Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics, *J. Phys. Chem. B*, **110**, 3533–9.
  53. Kulhanek, P., Bouchal, T., Durnik, I., et al. 2018, PMFLib - a toolkit for free energy calculations. <https://pmflib.ncbr.muni.cz>
  54. Humphrey, W., Dalke, A. and Schulten, K. 1996, VMD - visual molecular dynamics, *J. Mol. Graph.*, **14**, 33–8.

55. Blanchet, C., Pasi, M., Zakrzewska, K. and Lavery, R. 2011, CURVES plus web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures, *Nucleic Acids Res.*, **39**, W68–W73.
56. Lu, X.J. and Olson, W.K. 2003, 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, *Nucleic Acids Res.*, **31**, 5108–21.
57. Štěpán, J., Kulhánek, P., Lenčo, M., Strélcová, Z., Křenek, A. and Koča, J. 2013, Nemesi - a molecular modeling package, *J. Cheminformatics*, **5**, P12–P12.
58. Sherer, E.C., Harris, S.A., Soliva, R., Orozco, M. and Laughton, C.A. 1999, Molecular dynamics studies of DNA A-tract structure and flexibility, *J. Am. Chem. Soc.*, **121**, 5981–91.
59. Isaacs, R.J., Rayens, W.S. and Spielmann, H.P. 2002, Structural differences in the NOE-derived structure of G-T mismatched DNA relative to normal DNA are correlated with differences in C-13 relaxation-based internal dynamics, *J. Mol. Biol.*, **319**, 191–207.
60. La Rosa, G. and Zacharias, M. 2016, Global deformation facilitates flipping of damaged 8-oxo-guanine and guanine in DNA, *Nucleic Acids Res.*, **44**, 9591–9.
61. Lebbink, J.H.G., Fish, A., Reumer, A., Natrajan, G., Winterwerp, H.H.K. and Sixma, T.K. 2010, Magnesium coordination controls the molecular switch function of DNA mismatch repair protein MutS, *J. Biol. Chem.*, **285**, 13131–41.
62. Natrajan, G., Lamers, M.H., Enzlin, J.H., Winterwerp, H.H.K., Perrakis, A. and Sixma, T.K. 2003, Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates, *Nucleic Acids Res.*, **31**, 4814–21.
63. Fujii, S., Kono, H., Takenaka, S., Go, N. and Sarai, A. 2007, Sequence-dependent DNA deformability studied using molecular dynamics simulations, *Nucleic Acids Res.*, **35**, 6063–74.
64. Lankas, F. 2004, DNA sequence-dependent deformability - insights from computer simulations, *Biopolymers*, **73**, 327–39.
65. Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M. and Zhurkin, V.B. 1998, DNA sequence-dependent deformability deduced from protein-DNA crystal complexes, *Proc. Natl. Acad. Sci. USA*, **95**, 11163–8.
66. Stehlikova, K., Skalova, D., Zidkova, J., et al. 2014, Autosomal recessive limb-girdle muscular dystrophies in the Czech Republic, *BMC Neurol.*, **14**.
67. Tichy, L., Freiberger, T., Zapletalova, P., Soska, V., Ravcukova, B. and Fajkusova, L. 2012, The molecular basis of familial hypercholesterolemia in the Czech Republic: spectrum of LDLR mutations and genotype-phenotype correlations, *Atherosclerosis*, **223**, 401–8.
68. Reblova, K., Hrubá, Z., Prochazkova, D., Pazdirkova, R., Pouchla, S. and Fajkusova, L. 2013, Hyperphenylalaninemia in the Czech Republic: genotype-phenotype correlations and in silico analysis of novel missense mutations, *Clin. Chim. Acta*, **419**, 1–10.
69. Gojova, L., Jansova, E., Kulm, M., Pouchla, S. and Kozak, L. 2008, Genotyping microarray as a novel approach for the detection of ATP7B gene mutations in patients with Wilson disease, *Clin. Genet.*, **73**, 441–52.
70. Skalova, D., Zidkova, J., Vohanka, S., et al. 2013, CLCN1 Mutations in Czech patients with myotonia congenita, in silico analysis of novel and known mutations in the human dimeric skeletal muscle chloride channel, *Plos One*, **8**, e82549.
71. Grodecka, L., Hujova, P., Kramarek, M., et al. 2017, Systematic analysis of splicing defects in selected primary immunodeficiencies-related genes, *Clin. Immunol.*, **180**, 33–44.
72. Pigg, M., Bygum, A., Gånemo, A., et al. 2016, Spectrum of autosomal recessive congenital ichthyosis in scandinavia: clinical characteristics and novel and recurrent mutations in 132 patients, *Acta Derm. Venerol.*, **96**, 932–7.
73. Murayama, K., Tanaka, Y., Toda, T., Kashida, H. and Asanuma, H. 2013, Highly stable duplex formation by artificial nucleic acids acyclic threoninol nucleic acid (aTNA) and serinol nucleic acid (SNA) with acyclic scaffolds, *Chem. Eur. J.*, **19**, 14151–8.
74. Perez, A., Noy, A., Lankas, F., Luque, F.J. and Orozco, M. 2004, The relative flexibility of B-DNA and A-RNA duplexes: database analysis, *Nucleic Acids Res.*, **32**, 6144–51.
75. Faustino, I., Perez, A. and Orozco, M. 2010, Toward a consensus view of duplex RNA flexibility, *Biophys. J.*, **99**, 1876–85.